

The MDM advantage: Creating insight from big data



1

Introduction

2

**The importance
of understanding
your customers**

3

**How MDM
enhances big
data—and
vice versa**

4

**Leveraging
big data with
IBM InfoSphere
MDM**

5

**Unlock the
value of
big data**

6

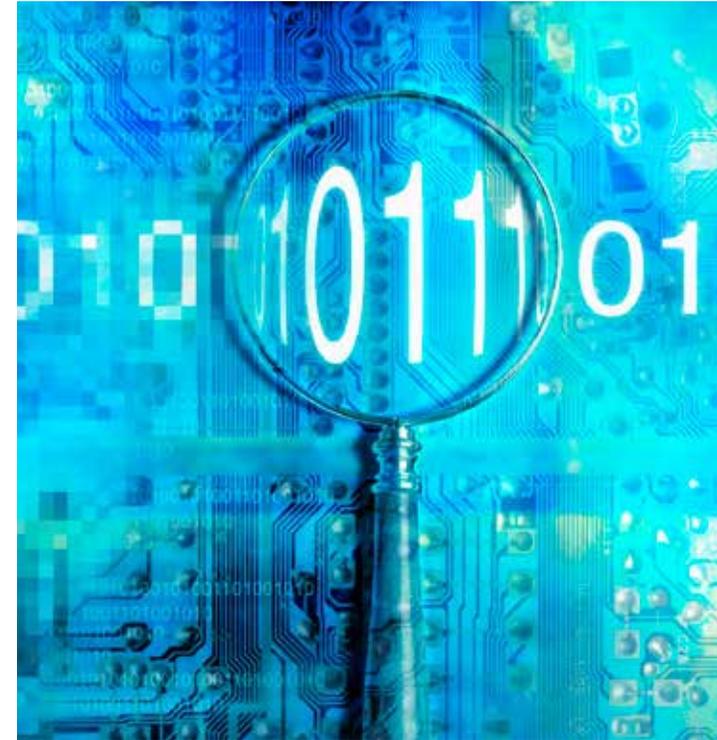
Why InfoSphere?

Introduction

Business leaders are eager to harness the power of big data. However, as the opportunity increases, ensuring that source information is trustworthy and protected becomes exponentially more difficult. If not addressed directly, end users may lose confidence in the insights generated from their data—which can result in a failure to act on opportunities or against threats.

Information integration and governance must be implemented within big data applications, providing appropriate governance and rapid integration from the start. By automating information integration and governance and employing it at the point of data creation, organizations can boost confidence in big data.

A solid information integration and governance program must become a natural part of big data projects, supporting automated discovery, profiling and understanding of diverse data sets to provide context and enable employees to make informed decisions. It must be agile to accommodate a wide variety of data and seamlessly integrate with diverse technologies, from data marts to Apache Hadoop systems. And it must automatically discover, protect and monitor sensitive information as part of big data applications.



To make the most of big data, you have to start with data you trust. But the sheer volume and complexity of big data means that the traditional manual methods of discovering, governing and correcting information are no longer feasible.

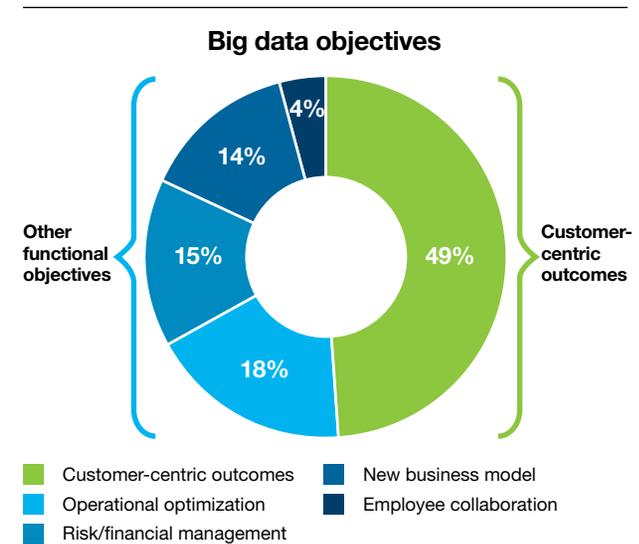
The importance of understanding your customers

When asked to rank their top three priorities for big data, nearly half of the respondents in a 2012 IBM Institute for Business Value study¹ named customer-centric objectives such as improving customer satisfaction and reducing churn as the top concern. This is not a new

trend; understanding today's "empowered consumer" was also identified as a high priority in both the 2011 IBM Global Chief Marketing Officer Study² and the 2012 IBM Global Chief Executive Officer Study.³

Companies clearly see big data as one key to understanding and predicting customer behavior. Transaction details, multichannel interactions, social media, syndicated data from sources such as loyalty cards, and other customer-related information are powerful new tools for creating a complete picture of customers' preferences and demands.

Armed with insights from big data, business leaders can make smarter decisions that improve the customer experience.



Top functional objectives identified by organizations with active big data pilots or implementations. Responses have been weighted and aggregated. Total respondents n = 1,061

Source: IBM Institute for Business Value. "Analytics: The Real-World Use of Big Data."

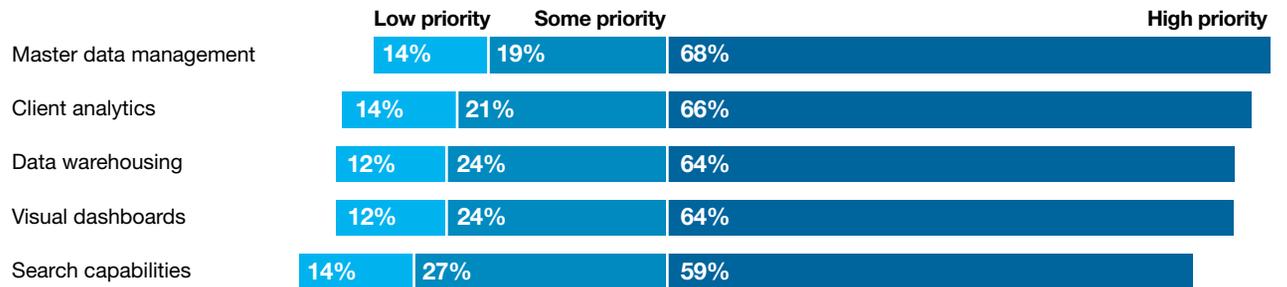
Top functional objectives identified by organizations with active big data pilots or implementations.

In another IBM Institute for Business Value study,⁴ CIOs identified the top five capabilities that enable translation of data into meaningful and useful intelligence:

1. Master data management (MDM)
2. Client analytics
3. Warehousing
4. Dashboards
5. Search

Executives expect these activities to lead to insights from both structured and unstructured data, and to enable them to understand customers at a level that is more individual, responsive and profound. Indeed, efforts to improve the customer experience should always start by knowing your customer.

Top five activities to turn data into intelligence



2011 CIO Study Q17: "What activities will you prioritize over the next 3 to 5 years to turn data into actionable information for your organization?" (n=3,018); some responses may add to over 100% due to rounding up.

Source: IBM Institute of Business Value. "Connect more: Intersecting insights from the IBM Global CEO, CMO and CIO Studies."

CIOs prioritize these five activities to turn data into intelligence.

More data, more understanding? Not necessarily

Gaining a full understanding and a single view of each customer—what makes them tick, why they buy, how they prefer to shop, why they switch, what they'll buy next, what factors lead them to recommend a company to others—is strategic for virtually every company.

Unfortunately, many organizations lack this understanding because they aren't taking advantage of traditional data sources. And if you don't know the customer based on the data you have now, it's going to be difficult to do so when you add big data to the mix. The complexity of using big data to gain additional insight boils down to several specific challenges:

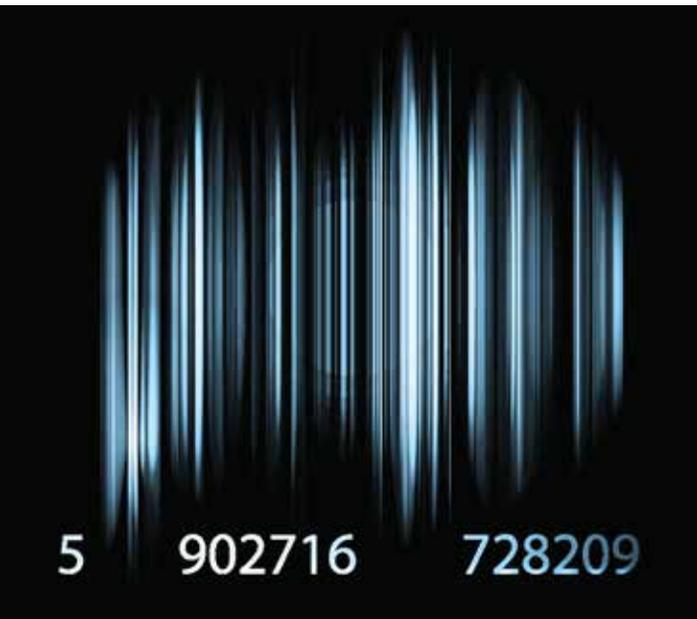


Big data extends beyond on-premise applications and data warehouses—

it may include, for example, tweets in which your customers discuss your products. Because of this, it can be harder to collect and analyze than traditional structured data. Potential insights can be buried in unstructured documents, spreadsheets, reports, email and social media.

The sheer volume of data about customers and products continues to skyrocket as well.

Exploiting this data to develop a full understanding of your customer can be daunting. How do you resolve hundreds of millions of records to satisfy the need for real-time insight or analyze large volumes of stored data to discover patterns and insights?



To complicate the situation further, **customer data is often captured in high-velocity streams.** Without real-time analysis capabilities, important insights may be lost.

CIOs are likely to find an MDM problem at the root of their big data challenges.

Information about customers and products is concealed within big data, and it must be linked to existing customer information that is already stored in traditional structured applications—otherwise, the new insights extracted from big data risk becoming yet another information silo.

MDM plays a crucial role in translating data into meaningful and useful intelligence. Together, big data and MDM can help CIOs and CMOs understand customers at an individual level and become more responsive to their needs. MDM creates context for big data by providing trusted information about how incoming unstructured data fits into the business environment. MDM can also help identify and structure big data that will be used in controlled environments. Conversely, big data creates context for MDM by providing new insights from social media and other sources, which helps companies build richer customer profiles.

How MDM enhances big data—and vice versa

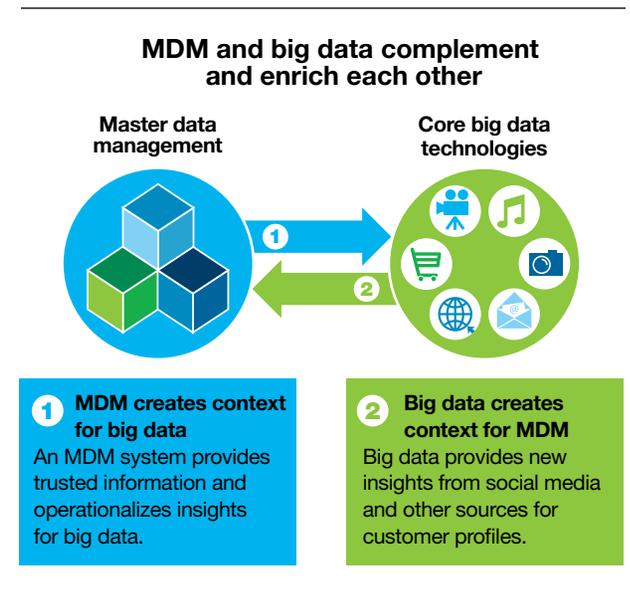
For years, master data management has helped companies start with the truth or facts about customers by defining the “golden record.” This golden record must be established before organizations can combine the power of MDM and big data to create further value.

To help enterprises create trusted insight as the volume, velocity and variety of data continue to explode, IBM offers several solutions designed to help organizations uncover previously unavailable insights and use them to support and inform decisions across the business. Combining the power of IBM® InfoSphere® MDM with IBM Watson™ Foundations creates a valuable

connection: big data technology can supply insights to MDM, and MDM can supply master data definitions to big data.

Analyzing big data at rest

By analyzing large volumes of stored data, organizations can discover patterns and insights that allow them to optimize processes and profitability. In environments where you quickly need to gain insight on hundreds of millions or even billions of records, the ability to match master data at the big-data scale is a must-have. Examples include reconciling large lead lists or matching citizens or customers against watch lists. Additionally, organizations want to analyze large volumes of stored data to discover patterns and insights.



MDM and big data complement and enrich each other.

Analyzing big data in motion

With certain kinds of data, there is no time to store it before acting on it because it is constantly changing. This data is often (but not always) generated by a sensor or other instrument. Speed matters for use cases such as fraud detection, emergent health care and public safety, where gaining insight into data in real time—and determining whether to act according to predefined rules—can be critical.

High-velocity, high-volume data calls for in-motion analytics. Streaming data may arrive linked to a master data identifier (a phone number, for example). In a neonatal ICU, sensor data is clearly matched to a particular infant. Financial

transactions are unambiguously matched to a credit card or Social Security number. However, not every piece of data that streams in is valuable—which is why it should be analyzed by a tool such as IBM InfoSphere Streams before being joined to master data.

InfoSphere Streams analyzes large data volumes with micro-latency. Rather than accumulating and storing data first, the software analyzes data as it flows in and identifies conditions that trigger alerts (such as outlier transactions that a bank flags as potentially fraudulent during the credit card authorization process). When this occurs, the data is passed out of the stream and matched with master data for better



business outcomes. InfoSphere Streams generates a summary of the insights derived from the stream analysis and matches it with trusted information, such as customer records, to augment the master data.

Analyzing a variety of big data

Perhaps the most widely recognized use case for big data involves analyzing social media or other highly variable data. If 50 percent of big data initiatives focus on improving the customer experience,⁵ these organizations must:

- 1) Improve the 360-degree view of the customer by adding in additional insight from social media
- 2) Discover additional relationship links based on insights from documents and unstructured text
- 3) Augment traditional product information with dynamically derived product traits based on web and social media feedback



MDM helps improve holistic views of the customer by enhancing master data with unstructured content. For example, Facebook postings may tell a hotel chain that a high proportion of its business guests have too many children for standard reward rooms. The hotel can then respond by providing reward privileges on larger suites for these high-value customers, potentially increasing their loyalty.

In addition, MDM helps organizations uncover relationship links that enhance existing knowledge about customers and open up opportunities to gain new business or save costs. For example, by establishing that two separate customers are part of the same household, the company can avoid sending duplicate mailings to the same address.

Leveraging big data with IBM InfoSphere MDM

Few organizations, if any, will ever process and analyze all the big data they accumulate; it will likely continue to grow and diversify ahead of attempts to map and analyze it. However, by identifying, classifying and using as much of the data as possible, organizations can gain valuable insights that would otherwise go undiscovered. IBM is focused on several important areas to help organizations make effective use of big data and improve the customer experience:

Enhancing the 360-degree view

Once an organization extracts insights from big data, it must link that new information to existing customer information in traditional structured applications.

Customer search —

Transaction history from MDM or ERP systems —

Products owned by the customer from MDM system —

Customer information from MDM system —

Unstructured internal information related to customer —

Third-party information related to customer —

Watson Explorer pulls in master data information to populate certain areas, presenting it along with information from other sources to create a complete view of an entity, such as a customer.



Watson Explorer is a platform for navigating and filtering large amounts of data in nearly any format. It offers a visual dashboard that joins master data about an individual with that person’s detailed transaction history from emails, social media, packaged applications and more. This data can help sales staff—particularly inbound and outbound marketers—connect with prospects on a more individual level and achieve better sales rates.

Designed to handle both social and sensor-generated data, Watson Explorer lets users specify an “axis” and then searches for

matching data that it can fuse onto existing master data. With master data as the starting point, Watson Explorer can also place real-time activity streams into context for analysis.

Intelligence, law enforcement and fraud analysts need to leverage all relevant pieces of information to connect the dots and make well-informed decisions instantly. Watson Explorer works from known master data to give intelligence agencies a unified view of all their information sources to deliver visibility, transparency and insight.

Correlating unstructured text to master records

Potential information about customers is often lost in unstructured documents, spreadsheets, reports, email and blogs—or it might be in enterprise content management systems and not available to MDM. With IBM InfoSphere MDM, organizations can correlate unstructured text to existing master records, discover linkages between text and relevant master entities, and enrich the master record with additional information.

These capabilities can be particularly useful in public safety, where MDM technologies can be used as an investigative tool.

Using MDM in fraud investigations

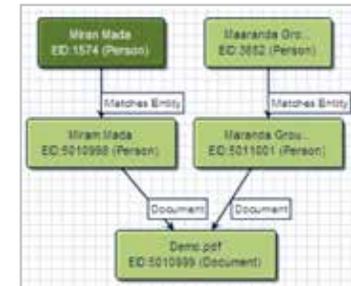
1. Potential suspect identified

Attribute	Value
Name	Miran Mada
Address	Via Riasc 21 Compine Italy
Citizenship	United States
Place of birth	Portland, Oregon
Date of birth	1938-09-15

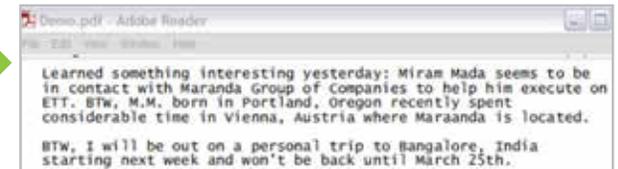
2. Search master records in InfoSphere MDM

Attribute	Code	Type
Address	ADDRESS	ADDRESS
Alternate name	AKA	NAME
Citizenship	CITIZENSHIP	MEMATTR
Date of birth	DOB	MEMDATE
Identifier	IDENTIFIER	IDENTIFIER
Name	NAME	NAME
Nationality	NATIONALITY	MEMATTR
Phone	PHONE	MEMPHONE
Place of birth	POB	MEMATTR
Remarks	REMARKS	TEXT

4. Matched entities identified; no automatic resolution of master record



3. Uncover links between entities from documents, emails, blogs and so on



Using MDM in fraud investigations.

Investigators want to know everything about a person from all sources, both structured and unstructured. After investigators establish a master record, they may later uncover links to that master record from documents, emails, blogs and so on. Since these sources tend to be less trustworthy than data from internal systems such as CRM or

One hospitality and travel organization used the bulk comparison process to match and consolidate 1 billion source records down to 250 million unique parties in just 22 hours.

ERP, the master record is not updated with this information, but instead is correlated so an investigator can quickly see relationships and connections between their structured record and unstructured content.

Another use case where big data can enhance master data involves product feedback. Using web and social media data, organizations can augment traditional product information to better understand customer sentiment. The resulting insights can help an organization update marketing offers, address technical issues, adjust product specs or change packaging. The company could also update or enhance the master record with insight from unstructured text.

Mastering information at big data scale

Enterprises are being flooded with data of all types, easily amassing terabytes and even petabytes of information as data volumes grow. That's why IBM designed InfoSphere MDM to meet the most demanding performance requirements.

InfoSphere MDM performs well not only in real-time transaction processing, but also in the initial bulk comparison and load process needed to establish the MDM hub. The bulk comparison and load process is scalable across processors and even across servers to help get the hub up and running quickly.

However, challenges can and do exist with the inherent limitations of entity resolution within relational databases. Identities must be compared against potentially all available entities, and to do this, comparison data must be retrieved from disk for processing. IBM InfoSphere BigInsights™—based on Hadoop—offers a solution for processing this data with near-limitless capacity that scales with CPU capacity.

Scalability and performance in resolving master data records can become an issue particularly when loading large lead lists and matching that list with known customers and prospects, or when matching records against large watch lists to combat threat and fraud. Analyzing large volumes of stored data enables organizations to

discover previously hidden patterns and insights that allow them to optimize processes and profitability.

For example, one healthcare insurance provider analyzed more than 400 million claims, looking for potentially deadly drug interactions. Text from patient phone calls is intricately nested and often mentions medications obtained elsewhere or without a claim form. A standard drug name might be Diflucan, but it could be called Flucanol, Batacan or yeast medicine in a transcript. The company’s internal system took more than 100 hours to perform this analysis.

However, when the company began using InfoSphere BigInsights to match transcripts to master data records, analysis time

dropped to just 10 hours for more than 6 TB of data.⁶ InfoSphere BigInsights also excels at extracting structured data from text analysis and matching it to master data. These capabilities pave the way for a new class of applications for customer experience analysis, social media analysis and fraud detection.

IBM has conducted performance benchmark tests in large deployments (over 100 million party records) and also has fully referenceable customer implementations of over 100 million party records scaling to hundreds of transactions per second.

Unlock the value of big data

InfoSphere MDM is a complete, flexible and proven MDM solution that creates trusted views to improve operational business processes, big data and analytics. Part of the InfoSphere platform, it supports all domains, architectural styles and use cases across all industries, and offers quick time-to-value through pre-built and customizable data models and business services.

Any enterprise that accumulates large volumes of fast-moving structured and unstructured data needs a well-defined approach to store, analyze and leverage it, as well as gauge its trustworthiness. InfoSphere MDM and IBM big data tools are keys that unlock the value of big data.



They help companies analyze big data in a timely manner, match it to master data and derive actionable information to inform decisions and extract insights. InfoSphere MDM complements big-data analytics

tools and helps organizations deliver trusted information to win sales, satisfy and retain customers, improve operations and increase compliance.

Why InfoSphere?

As a critical element of Watson Foundations, the IBM big data and analytics platform, InfoSphere Information Integration and Governance (IIG) provides optimal scalability and performance for massive data volumes, agile and rightsized integration and governance for the increasing velocity of data, and support and protection for a wide variety of data types and big data systems. InfoSphere helps make big data and analytics projects successful by delivering business users the confidence to act on insight. InfoSphere capabilities include:

- **Metadata, business glossary and policy management:** Define metadata, business terminology and governance policies with IBM InfoSphere Business Information Exchange.
- **Data integration:** Handle all integration requirements, including batch data transformation and movement (IBM InfoSphere Information Server), real-time replication (IBM InfoSphere Data Replication) and data federation (IBM InfoSphere Federation Server).
- **Data quality:** Parse, standardize, validate and match enterprise data with IBM InfoSphere Information Server for Data Quality.
- **Master data management:** Act on a trusted view of your customers, products, suppliers, locations and accounts with InfoSphere MDM.
- **Data lifecycle management:** Manage the data lifecycle from test data creation through retirement and archiving with IBM InfoSphere Optim™.
- **Data security and privacy:** Continuously monitor data access and protect repositories from data breaches, and support compliance with IBM InfoSphere Guardium®. Mask and help protect sensitive data with InfoSphere Optim.

Additional resources

To learn more about the IBM approach to information integration, MDM and governance for big data, please contact your IBM representative or IBM Business Partner, or visit:

- ibm.com/software/data/information-integration-governance
- ibm.com/software/products/us/en/infosphere-master-data-management



© Copyright IBM Corporation 2014

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
March 2014

IBM, the IBM logo, ibm.com, BigInsights, Guardium, IBM Watson, InfoSphere, and Optim are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at ibm.com/legal/copytrade.shtml

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data discussed herein is presented as derived under specific operating conditions. Actual results may vary.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

Actual available storage capacity may be reported for both uncompressed and compressed data and will vary and may be less than stated.



Please Recycle

^{1,5} “Analytics: The real-world use of big data.” ibm.com/services/us/gbs/thoughtleadership/ibv-big-data-at-work.html

² 2011 IBM Global Chief Marketing Officer Study. ibm.com/services/us/cmo/cmoustudy2011/cmo-registration.html

³ 2012 IBM Global Chief Executive Officer Study. ibm.com/services/us/en/c-suite/ceostudy2012

⁴ IBM Institute of Business Value. “Connect more: Intersecting insights from the IBM Global CEO, CMO and CIO Studies.” ibm.com/services/us/gbs/thoughtleadership/connectmore

⁶ DeRoos, Dirk; Deutsch, Tom; Eaton, Chris; Lapis, George; Zikopoulos, Paul. “Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data.” McGraw-Hill, 2012. ibm.com/common/ssi/cgi-bin/ssialias?subtype=WH&infotype=SA&appname=SWGE_IM_DD_USEN&htmlfid=IML14297USEN&attachment=IML14297USEN.PDF